

# A randomized inertial primal-dual fixed point algorithm for monotone inclusions

Meng Wen<sup>1,2</sup>, Shigang Yue<sup>4</sup>, Yuchao Tang<sup>3</sup>, Jigen Peng<sup>1,2</sup>

1. School of Mathematics and Statistics, Xi'an Jiaotong University, Xi'an 710049, P.R. China
2. Beijing Center for Mathematics and Information Interdisciplinary Sciences, Beijing, P.R. China
3. Department of Mathematics, NanChang University, Nanchang 330031, P.R. China
4. School of Computer Science, University of Lincoln, LN6 7TS, UK

**Abstract** In this paper, we propose a randomized inertial block-coordinate primal-dual fixed point algorithm to solve a wide array of monotone inclusion problems based on the modification of the heavy ball method of Nesterov [41]. These methods rely on a sweep of blocks of variables which are activated at each iteration according to a random rule. To this end we formulate the inertial version of the Krasnosel'skii-Mann algorithm for approximating the set of fixed points of a quasinonexpansive operator, for which we also provide an exhaustive convergence analysis. As a by-product, we can obtain some inertial block-coordinate operator splitting methods for solving composite monotone inclusion and convex minimization problems.

**Keywords:** preconditioning; block-coordinate algorithm; inertial algorithm; stochastic quasi-Fejér sequence

**MR(2000) Subject Classification** 47H09, 90C25,

## 1 Introduction

The problem of approaching the set of zeros of a sum of monotone operators or minimizing a sum of proper lower-semicontinuous convex functions by means of primal-dual algorithms, where various linear operators are involved in the formulation, solving

---

\* Corresponding author.

E-mail address: jgpengxjtu@126.com

jointly its primal and dual forms and none of the linear operators needs to be inverted, continues to be a very attractive research area. This is due to its avoiding such inversions may offer a significant advantage in terms of computational complexity when dealing with large-scale problems [20-34] and its applicability in the context of solving real-life problems which can be modeled as nondifferentiable convex optimization problems, like those arising in image processing, signal recovery, support vector machines classification, location theory, clustering, network communications, etc.

The main advantage of inertial methods is it can change greatly improves the performance of the scheme. These benefits have long been recognized [18, 35-37] and have become increasingly important in many monotone inclusion problems. In instances, Radu Ioan, Ernő Robert and Christopher [1] considered an inertial Douglas-Rachford splitting for finding the set of zeros of the sum of two maximally monotone operators in Hilbert spaces and investigate its convergence.

Recently, Patrick L. and Jean-Christophe [2] introduced a block-coordinate fixed point algorithms with applications to nonlinear analysis and optimization in Hilbert spaces based on a notion of stochastic quasi-Fejér monotonicity. The algorithms were composed of quasinonexpansive operators or compositions of averaged nonexpansive operators. In addition, they proved weak and strong convergence results of the sequences generated by these algorithms.

Motivated and inspired by the above results, we introduce a randomized inertial block-coordinate primal-dual fixed point algorithm to solve monotone inclusion problems. The main benefit of block-coordinate algorithms is to result in implementations with reduced complexity and memory requirements per iteration (see [38-40]). We obtain the weak and strong convergence theorem of proposed algorithms. Furthermore, we using our algorithms to solve some monotone inclusion problems and large-scale convex minimization problems. Our results also extend and improve the corresponding results of Patrick L. and Jean-Christophe [2] and Jean-Christophe and Audrey [3] and many others.

The rest of this paper is organized as follows. In the next section, we recall the conception of stochastic quasi-Fejér monotone and some related notations and then deduce the idea we proposed stochastic inertial quasi-Fejér monotonicity. Furthermore, we prove almost sure convergence results for a randomized inertial iteration scheme.

In section 3, we using this scheme to design a stochastic inertial block-coordinate fixed point algorithms for relaxed iterations of quasinonexpansive operators and a stochastic inertial block-coordinate algorithms involving compositions of averaged nonexpansive operators. In section 4, Based on the algorithm in Section 3, we present a preconditioned stochastic inertial block-coordinate forward-backward algorithm. In the final section, we consider the application of the presented algorithms. First, we give a novel inertial block-coordinate prima-dual algorithms for solving a zero of a sum of monotone operators. Moreover, we prove the convergence of proposed algorithms. Second, an inertial block-coordinate primal-dual algorithms are designed to solve composite convex optimization problems, and we study their convergence.

## 2 Stochastic inertial quasi-Fejér monotonicity

In various areas of nonlinear analysis and optimization to unify the convergence proofs of deterministic algorithms, Fejér monotonicity has been exploited far-ranging see, e.g., [21-24]. In the late 1960s, this conception was reconsidered in a stochastic setting in Euclidean spaces [25-27]. In this section, we introduce a conception of stochastic inertial quasi-Fejér monotone sequence in Hilbert spaces and use the results to a general stochastic inertial iterative method. We will use the following notation throughout the paper.

**Notation 2.1.** Let  $H$  be a separable real Hilbert space with inner product  $\langle \cdot, \cdot \rangle$ , norm  $\|\cdot\|$  and Borel  $\sigma$ -algebra  $\mathcal{B}$ . In  $H$ , we write  $\rightarrow$  and  $\rightharpoonup$  indicate respectively weak and strong convergence. The sets of strong and weak cluster point of a sequence  $(x_n)_{n \in \mathbb{N}}$  in  $H$  are recorded as  $\mathfrak{Q}(x_n)_{n \in \mathbb{N}}$  and  $\mathfrak{Y}(x_n)_{n \in \mathbb{N}}$  respectively.  $(\Omega, \mathcal{F}, P)$  is the underlying probability space. A  $H$ -valued random variable is a measurable map  $x : (\Omega, \mathcal{F}) \rightarrow (H, \mathcal{B})$ . The smallest  $\sigma$ -algebra generated by a family  $\Phi$  of random variables is denoted by  $\sigma(\Phi)$ . The expectation is denoted by  $E(\cdot)$ . Let  $\mathfrak{F} = (\mathcal{F})_{n \in \mathbb{N}}$  be a sequence of sub-sigma algebras of  $\mathcal{F}$  such that  $(\forall n \in \mathbb{N}) \mathcal{F}_n \subset \mathcal{F}_{n+1}$ . We denote by  $\ell_+(\mathfrak{F})$  the set of sequences of  $[0, +\infty)$ -valued random variables  $(\pi_n)_{n \in \mathbb{N}}$  such that, for every  $n \in \mathbb{N}$ ,  $\pi_n$  is  $\mathcal{F}_n$ -measurable. We set

$$(\forall p \in (0, +\infty)) \quad \ell_+^p(\mathfrak{F}) = \{(\pi_n)_{n \in \mathbb{N}} \in \ell_+(\mathfrak{F}) \mid \sum_{n \in \mathbb{N}} \pi_n^p < +\infty \quad P - a.s.\},$$

and

$$\ell_+^\infty(\mathfrak{F}) = \{(\pi_n)_{n \in \mathbb{N}} \in \ell_+(\mathfrak{F}) \mid \sup_{n \in \mathbb{N}} \pi_n < +\infty \text{ } P\text{-a.s.}\}.$$

Let  $(x_n)_{n \in \mathbb{N}}$  be sequences of  $H$ -valued random variables, we write for  $\mathfrak{X} = (\mathcal{X}_n)_{n \in \mathbb{N}}$ , where  $(\forall n \in \mathbb{N}) \mathcal{X}_n = \sigma(x_0, \dots, x_n)$ .

**Lemma 2.1.** ([21, Corollary 2.14]). For  $\forall a \in \mathbb{R}$  and  $\forall x, y \in H$ ,

$$\|ax + (1-a)y\|^2 = a\|x\|^2 + (1-a)\|y\|^2 - a(1-a)\|x-y\|^2. \quad (2.1)$$

**Lemma 2.2.** ([2]). Let  $D$  be a nonempty closed subset of  $H$ , let  $\Phi : [0, +\infty) \rightarrow [0, +\infty)$  be a strictly increasing function such that  $\lim_{t \rightarrow +\infty} \Phi(t) = +\infty$ , and let  $(x_n)_{n \in \mathbb{N}}$  be a sequence of  $H$ -valued random variables. Suppose that, for every  $z \in D$ , there exist  $(\zeta_n(z))_{n \in \mathbb{N}} \in \ell_+^1(\mathfrak{X})$ ,  $(\xi_n(z))_{n \in \mathbb{N}} \in \ell_+(\mathfrak{X})$ , and  $(\theta_n(z))_{n \in \mathbb{N}} \in \ell_+^1(\mathfrak{X})$  such that the following is satisfied  $P$ -a.s.:

$$E(\Psi(\|x_{n+1} - z\|) | \mathcal{X}_n) + \zeta_n(z) \leq (1 + \xi_n(z))\Psi(\|x_n - z\|) + \theta_n(z). \quad (2.2)$$

Then the following hold:

- (i)  $\sum_{n \in \mathbb{N}} \xi_n(z) < +\infty$   $P$ -a.s.,  $\forall z \in D$ .
- (ii)  $(x_n)_{n \in \mathbb{N}}$  is bounded  $P$ -a.s.
- (iii) There exists  $\bar{\Omega} \in \mathcal{F}$  such that  $P(\bar{\Omega}) = 1$  and, for every  $\omega \in \bar{\Omega}$  and every  $z \in D$ ,  $(\|x_n(\omega) - z\|)_{n \in \mathbb{N}}$  converges.
- (iv) Suppose that  $\mathfrak{Q}(x_n)_{n \in \mathbb{N}} \subset D$   $P$ -a.s. Then  $(x_n)_{n \in \mathbb{N}}$  converges weakly  $P$ -a.s. to a  $D$ -valued random variable.
- (v) Suppose that  $\mathfrak{Y}(x_n)_{n \in \mathbb{N}} \cap D \neq \emptyset$   $P$ -a.s. Then  $(x_n)_{n \in \mathbb{N}}$  converges strongly  $P$ -a.s. to a  $D$ -valued random variable.
- (vi) Suppose that  $\mathfrak{Y}(x_n)_{n \in \mathbb{N}} \neq \emptyset$   $P$ -a.s. and that  $\mathfrak{Q}(x_n)_{n \in \mathbb{N}} \subset D$   $P$ -a.s. Then  $(x_n)_{n \in \mathbb{N}}$  converges strongly  $P$ -a.s. to a  $D$ -valued random variable.

**Lemma 2.3.** (Opial). Let  $C$  be a nonempty set of  $H$  and  $(x_n)_{n \in \mathbb{N}}$  be a sequence in  $H$  such that the following two conditions hold:

- (a) for every  $x \in C$ ;  $\lim_{n \rightarrow +\infty} \|x_n - x\|$  exists.
- (b) every sequential weak cluster point of  $(x_n)_{n \in \mathbb{N}}$  is in  $C$ .

Then  $(x_n)_{n \in \mathbb{N}}$  converges weakly to a point in  $C$ .

**Lemma 2.4.** ([20, Theorem 3]). We denote by  $\mathcal{A}(H, \beta)$  the set of  $\beta$ -averaged operators on  $H$ . Let  $\beta_1 \in (0, 1)$ ,  $\beta_2 \in (0, 1)$ ,  $T_1 \in \mathcal{A}(H, \beta_1)$ , and  $T_2 \in \mathcal{A}(H, \beta_2)$ . Then  $T_1 \circ T_2 \in \mathcal{A}(H, \beta')$ , where

$$\beta' = \frac{\beta_1 + \beta_2 - 2\beta_1\beta_2}{1 - \beta_1\beta_2}.$$

**Theorem 2.1.** Let  $C$  be a nonempty closed affine subset of  $H$ , let  $(\lambda_n)_{n \in \mathbb{N}}$  be a sequence in  $(0, 1]$ , and let  $(t_n)_{n \in \mathbb{N}}$  and  $(x_n)_{n \in \mathbb{N}}$  be sequences of  $C$ -valued random variables. Suppose that the following hold:

- (i)  $w_n = x_n + \alpha_n(x_n - x_{n-1})$ ,  $x_{n+1} = w_n + \lambda_n(t_n - w_n)$ ,  $\forall n \in \mathbb{N}$ .
- (ii)  $x_0, x_1$  are arbitrarily chosen in  $C$ ,  $(\alpha_n)_{n \geq 1}$  is nondecreasing with  $\alpha_1 = 0$  and  $0 \leq \alpha_n \leq \alpha < 1$ ,  $\forall n \geq 1$  and  $\lambda, \tau, \delta > 0$  are such that  $\delta > \frac{\alpha^2(1+\alpha)+\alpha\tau}{1-\alpha^2}$  and  $0 < \lambda \leq \lambda_n \leq \frac{\delta-\alpha[\alpha(1+\alpha)+\alpha\delta+\tau]}{\delta[1+\alpha(1+\alpha)+\alpha\delta+\tau]}$ ,  $\forall n \geq 1$ .
- (iii)  $\forall n \in \mathbb{N}$ ,  $E(\|t_n - y\|^2 | \mathcal{X}_n) \leq \|w_n - y\|^2$  is satisfied  $P$ -a.s.

Then

$$\sum_{n \in \mathbb{N}} E(\|t_n - w_n\|^2 | \mathcal{X}_n) < +\infty \text{ } P\text{-a.s.}$$

Moreover, assuming that:

- (iv)  $\mathfrak{Q}(x_n)_{n \in \mathbb{N}} \subset C$   $P$ -a.s. Then  $(x_n)_{n \in \mathbb{N}}$  converges weakly  $P$ -a.s. to a  $C$ -valued random variable  $\hat{x}$ . Furthermore, if
  - (v)  $\mathfrak{Y}(x_n)_{n \in \mathbb{N}} \neq \emptyset$   $P$ -a.s.,
- then  $(x_n)_{n \in \mathbb{N}}$  converges strongly  $P$ -a.s. to  $\hat{x}$ .

*Proof.* We proceed with the following steps.

Step 1. Because of the choice of  $\delta$ ,  $\lambda_n \in (0, 1) \forall n \geq 1$ . Moreover, since  $C$  is affine, we can know that the iterative scheme provides a well-defined sequence in  $C$ . Let  $y \in C$  and  $n \geq 1$ . From (2.1) and (iii), we can know that

$$\begin{aligned} \|x_{n+1} - y\|^2 &= \|(1 - \lambda_n)w_n + \lambda_n t_n - y\|^2 \\ &= \|(1 - \lambda_n)(w_n - y) + \lambda_n(t_n - y)\|^2 \\ &= (1 - \lambda_n)\|w_n - y\|^2 + \lambda_n\|t_n - y\|^2 - \lambda_n(1 - \lambda_n)\|t_n - w_n\|^2 \\ &\leq \|w_n - y\|^2 - \lambda_n(1 - \lambda_n)\|t_n - w_n\|^2. \end{aligned} \tag{2.3}$$

Using (2.1) again, we obtain

$$\|w_n - y\|^2 = \|(1 + \alpha_n)(x_n - y) - \alpha_n(x_{n-1} - y)\|^2$$

$$= (1 + \alpha_n)\|x_n - y\|^2 - \alpha_n\|x_{n-1} - y\|^2 + \alpha_n(1 + \alpha_n)\|x_n - x_{n-1}\|^2. \quad (2.4)$$

So from (2.3), we have

$$\begin{aligned} (\forall n \geq 1) & \mathbb{E}(\|x_{n+1} - y\|^2 | \mathcal{X}_n) - (1 + \alpha_n)\|x_n - y\|^2 + \alpha_n\|x_{n-1} - y\|^2 \\ & \leq -\lambda_n(1 - \lambda_n)\mathbb{E}(\|t_n - w_n\|^2 | \mathcal{X}_n) + \alpha_n(1 + \alpha_n)\|x_n - x_{n-1}\|^2. \end{aligned} \quad (2.5)$$

Furthermore, we have

$$\begin{aligned} \mathbb{E}(\|t_n - w_n\|^2 | \mathcal{X}_n) &= \left\| \frac{1}{\lambda_n}(x_{n+1} - x_n) + \frac{\alpha_n}{\lambda_n}(x_{n-1} - x_n) \right\|^2 \\ &= \frac{1}{\lambda_n^2}\|x_{n+1} - x_n\|^2 + \frac{\alpha_n^2}{\lambda_n^2}\|x_{n-1} - x_n\|^2 + 2\frac{\alpha_n}{\lambda_n^2}\langle x_{n+1} - x_n, x_{n-1} - x_n \rangle \\ &\geq \frac{1}{\lambda_n^2}\|x_{n+1} - x_n\|^2 + \frac{\alpha_n^2}{\lambda_n^2}\|x_{n-1} - x_n\|^2 \\ &\quad + \frac{\alpha_n}{\lambda_n^2}(-v_n\|x_{n+1} - x_n\|^2 - \frac{1}{v_n}\|x_{n-1} - x_n\|^2) \\ &= \frac{1 - \alpha_n v_n}{\lambda_n^2}\|x_{n+1} - x_n\|^2 + \frac{\alpha_n^2 v_n - \alpha_n}{\lambda_n^2 v_n}\|x_{n-1} - x_n\|^2, \end{aligned} \quad (2.6)$$

where  $v_n = \frac{1}{\alpha_n + \delta \lambda_n}$ .

Put (2.6) into (2.5), we have

$$\begin{aligned} & \mathbb{E}(\|x_{n+1} - y\|^2 | \mathcal{X}_n) - (1 + \alpha_n)\|x_n - y\|^2 + \alpha_n\|x_{n-1} - y\|^2 \\ & \leq \frac{(1 - \lambda_n)(\alpha_n v_n - 1)}{\lambda_n}\|x_{n+1} - x_n\|^2 + \mu_n\|x_n - x_{n-1}\|^2, \end{aligned} \quad (2.7)$$

where

$$\mu_n = \alpha_n(1 + \alpha_n) + \alpha_n(1 - \lambda_n)\frac{1 - v_n \alpha_n}{v_n \lambda_n} > 0, \quad (2.8)$$

since  $v_n \alpha_n < 1$  and  $\lambda_n \in (0, 1)$ .

Now, we consider the choice of  $v_n$  and set

$$\delta = \frac{1 - v_n \alpha_n}{v_n \lambda_n}.$$

Therefore, from (2.8) we can know that

$$\mu_n = \alpha_n(1 + \alpha_n) + \alpha_n(1 - \lambda_n)\delta \leq \alpha(1 + \alpha) + \alpha\delta \quad \forall n \geq 1. \quad (2.9)$$

In the following, we set  $\psi_n = \|x_n - y\|^2$ ,  $\forall n \in \mathbb{N}$  and  $\eta_{n+1} = E(\psi_{n+1}|\mathcal{X}_n) - \alpha_{n+1}\psi_n + \mu_{n+1}\|x_{n+1} - x_n\|^2$ ,  $\forall n \geq 1$ . With the same proof of [1], we can obtain that

$$\sum_{n \in \mathbb{N}} \|x_{n+1} - x_n\|^2 < +\infty. \quad (2.10)$$

We have proven above that for an arbitrary  $y \in C$  the inequality (2.7) is true. By Step 1, (2.9) and Lemma 2.2 we derive that exists  $\Omega \in C$  such that  $P(\Omega) = 1$  and, for every  $\omega \in \Omega$  and every  $y \in C$   $\lim_{n \rightarrow \infty} \|x_n(\omega) - y\|$  exists (in (2.7), we still consider that  $v_n \alpha_n < 1$ ,  $\forall n \geq 1$ ). On the other hand, from (i) we have

$$\begin{aligned} (\forall n \geq 1) E(\|t_n - w_n\||\mathcal{X}_n) &= \frac{1}{\lambda_n} E(\|x_{n+1} - w_n\||\mathcal{X}_n) \\ &\leq \frac{1}{\lambda} E(\|x_{n+1} - w_n\||\mathcal{X}_n) \\ &\leq \frac{1}{\lambda} (\|x_{n+1} - x_n\| + \alpha \|x_n - x_{n-1}\|), \end{aligned} \quad (2.11)$$

therefore, by (2.10) we have

$$\sum_{n \in \mathbb{N}} E(\|t_n - w_n\|^2|\mathcal{X}_n) < +\infty \text{ P-a.s.}$$

Step 2. We will prove the convergence of  $(x_n)_{n \in \mathbb{N}}$ .

By the definition of  $w_n$ , we can know that

$$\begin{aligned} \|w_n - y\|^2 &= \|(x_n - y) - \alpha_n(x_n - x_{n-1})\|^2 \\ &\leq \|x_n - y\|^2 + \alpha_n^2 \|x_n - x_{n-1}\|^2 + 2\alpha_n \|x_n - y\| \|x_n - x_{n-1}\|. \end{aligned} \quad (2.12)$$

Put (2.12) into (2.3), we have

$$\begin{aligned} \|x_{n+1} - y\|^2 &\leq \|x_n - y\|^2 + \alpha_n^2 \|x_n - x_{n-1}\|^2 + 2\alpha_n \|x_n - y\| \|x_n - x_{n-1}\| \\ &\quad - \lambda_n(1 - \lambda_n) \|t_n - w_n\|^2. \end{aligned} \quad (2.13)$$

Set

$$\xi_n(y) = \lambda_n(1 - \lambda_n) \|t_n - w_n\|^2$$

and

$$\theta_n(y) = \alpha_n^2 \|x_n - x_{n-1}\|^2 + 2\alpha_n \|x_n - y\| \|x_n - x_{n-1}\|.$$

So, from (iv) and Lemma 2.2 (iv) applied with  $\Psi : t \rightarrow t^2$ , we obtain that  $(x_n)_{n \in \mathbb{N}}$  converges weakly P-a.s. to a  $C$ -valued random variable  $\hat{x}$ . Likewise, from (iv)-(v)

and Lemma 2.2 (vi) applied with  $\Psi : t \rightarrow t^2$ , we obtain the strong convergence of  $(x_n)_{n \in \mathbb{N}}$ .  $\square$

**Definition 2.2.** An operator  $T : H \rightarrow H$  is nonexpansive, if the following inequality holds for any  $x, y \in H$ :

$$\|Tx - Ty\| \leq \|x - y\|.$$

A point  $x \in H$  is a fixed point of  $T$  provided  $Tx = x$ . Denote by  $\text{Fix}(T)$  the set of fixed points of  $T$ , that is,  $\text{Fix}(T) = \{x \in H, Tx = x\}$ . Let  $(x_n)_{n \in \mathbb{N}}$  be a sequence in  $H$  and  $z \in H$  such that  $x_n \rightharpoonup z$  and  $Tx_n - x_n \rightarrow 0$  as  $n \rightarrow +\infty$ . Then  $z \in \text{Fix}(T)$ . This is called  $T$  is demicompact at  $z \in H$  [21, Corollary 4.18].

By Theorem 2.1, we can obtain the following Corollary.

**Corollary 2.3.** Let  $T : H \rightarrow H$  be a nonexpansive operator such that  $\text{Fix}(T) \neq \emptyset$ , and  $(\lambda_n)_{n \in \mathbb{N}}$  be a sequence in  $[0, 1]$ . Let  $x_0, x_1$  be  $H$ -valued random variables which are chosen arbitrarily. For  $n \geq 0$ :

$$\begin{cases} w_n = x_n + \alpha_n(x_n - x_{n-1}), \\ x_{n+1} = x_n + \lambda_n(Tx_n - x_n). \end{cases}$$

Then the following hold:

- (i)  $(x_n)_{n \in \mathbb{N}}$  converges weakly P-a.s. to a  $\text{Fix}(T)$ -valued random variable.
- (ii) Suppose that  $T$  is demicompact at 0 (see Definition 2.1). Then  $(x_n)_{n \in \mathbb{N}}$  converges strongly P-a.s. to a  $\text{Fix}(T)$ -valued random variable.

*Proof.* Set  $D = \text{Fix}(T)$ . Since  $T$  is continuous,  $T$  is measurable and  $D$  is closed. Now let  $y \in D$  and set  $(n \in \mathbb{N})$ ,  $t_n = Tw_n$ . Then, by the nonexpansiveness of  $T$ , we have for every  $n \in \mathbb{N}$

$$\begin{cases} x_{n+1} = w_n + \lambda_n(t_n - w_n), \\ \sum_{n \in \mathbb{N}} E(\|t_n - w_n\|^2 | \mathcal{X}_n) = \|Tw_n - w_n\|^2, \\ \sum_{n \in \mathbb{N}} E(\|t_n - y\|^2 | \mathcal{X}_n) = \|Tw_n - Ty\|^2 \leq \|w_n - y\|^2. \end{cases} \quad (2.14)$$

It is satisfied properties (i)-(iii) of Theorem 2.1. Therefore, from (2.14) and the conclusion of (iii) in Theorem 2.1, we know that exists  $\bar{\Omega} \in \mathcal{F}$  such that  $P(\bar{\Omega}) = 1$  and

$$\sum_{n \in \mathbb{N}} \|Tw_n(\omega) - w_n(\omega)\|^2 < +\infty, \forall \omega \in \bar{\Omega}. \quad (2.15)$$



With the same proof of Step1 in Theorem 2.1, we can know that for every  $y \in C$ ,  $\forall \omega \in \bar{\Omega}$ ,  $\lim_{n \rightarrow \infty} \|x_n(\omega) - y\|$  exists. On the other hand, let  $x$  be a sequential weak cluster point of  $(x_n)_{n \in \mathbb{N}}$ , that is, the latter has a subsequence  $(x_{n_k})_{k \in \mathbb{N}}$  such that  $x_{n_k} \rightharpoonup x$  as  $k \rightarrow \infty$ . By (2.10) and the definition of  $w_n$  and the upper bound for  $\alpha_n$ , we get  $w_{n_k} \rightharpoonup x$ . Then, by (2.15) we know there exists  $\hat{\Omega} \subset \bar{\Omega}$  such that  $\hat{\Omega} \in \mathcal{F}$  such that  $P(\hat{\Omega}) = 1$  and, for every  $\omega \in \hat{\Omega}$   $Tw_{n_k}(\omega) - w_{n_k}(\omega) \rightarrow 0$  as  $k \rightarrow \infty$ . Applying now Definition 2.1 for the sequence  $(w_{n_k}(\omega))_{k \in \mathbb{N}}$  we conclude that  $x \in \text{Fix}(T)$ . Since the two assumptions of Lemma 2.3 are verified, it follows that  $(x_n(\omega))_{n \in \mathbb{N}}$  converges weakly to a point in  $\text{Fix}(T)$ . □

### 3 A stochastic inertial block-coordinate fixed point algorithms

**Notation 3.1.** Let  $H_1, \dots, H_m$  are separable real Hilbert spaces and  $\mathbf{H} = H_1 \oplus \dots \oplus H_m$  be their direct Hilbert sum. The inner products and norms of these spaces are all denoted by  $\langle \cdot, \cdot \rangle$  and  $\|\cdot\|$ , and  $\mathbf{x} = (x_1, \dots, x_m)$  indicates a generic vector in  $\mathbf{H}$ . Let  $(x_n)_{n \in \mathbb{N}}$  be a sequence of  $H$ -valued random variables, we set  $\mathcal{X}_n = \sigma(x_0, \dots, x_n)$ .

**Definition 3.1.** ([21]) An operator  $\mathbf{T} : \mathbf{H} \rightarrow \mathbf{H}$  is *quasinonexpansive* if

$$\|\mathbf{T}\mathbf{x} - \mathbf{z}\| \leq \|\mathbf{x} - \mathbf{z}\|, \forall \mathbf{z} \in \text{Fix}(\mathbf{T}), \forall \mathbf{x} \in \mathbf{H}. \quad (3.1)$$

**Theorem 3.2.** Let  $(\forall n \in \mathbb{N}) \mathbf{T}_n : \mathbf{H} \rightarrow \mathbf{H} : \mathbf{x} \mapsto (T_{i,n}\mathbf{x})_{1 \leq i \leq m}$  be a quasinonexpansive operator where, for every  $i \in \{1, \dots, m\}$ ,  $T_{i,n} : \mathbf{H} \rightarrow H_i$  is measurable. Suppose that  $(\lambda_n)_{n \in \mathbb{N}}$  is a sequence in  $(0, 1)$ , and set  $M = \{0, 1\}^m \setminus \{\mathbf{0}\}$ . Let  $\mathbf{x}_0, \mathbf{x}_1$  be  $\mathbf{H}$ -valued random variables which are arbitrarily chosen, and let  $(\varepsilon_n)_{n \in \mathbb{N}}$  be identically distributed  $M$ -valued random variables. For  $n \geq 0$ ,

$$\begin{cases} \mathbf{w}_n = \mathbf{x}_n + \alpha_n(\mathbf{x}_n - \mathbf{x}_{n-1}), \\ x_{i,n+1} = w_{i,n} + \varepsilon_{i,n}\lambda_n(T_{i,n}(w_{1,n}, \dots, w_{m,n}) - w_{i,n}), i = \{1, \dots, m\}, \end{cases} \quad (3.2)$$

and set  $\mathcal{E}_n = \sigma(\varepsilon_n)$ . Moreover, suppose that the following hold:

(i)  $\mathbf{D} = \bigcap_{n \in \mathbb{N}} \text{Fix}(\mathbf{T}_n) \neq \emptyset$ .

(ii) For every  $n \in \mathbb{N}$ ,  $\mathcal{E}_n$  and  $\mathcal{X}_n$  are independent.

(iii)  $p_i = P[\varepsilon_{i,0} = 1] > 0$ ,  $\forall i \in \{1, \dots, m\}$ .

(iv)  $(\alpha_n)_{n \geq 1}$  is nondecreasing with  $\alpha_1 = 0$  and  $0 \leq \alpha_n \leq \alpha < 1$ ,  $\forall n \geq 1$  and  $\lambda, \tau, \delta > 0$  are such that  $\delta > \frac{\alpha^2(1+\alpha)+\alpha\tau}{1-\alpha^2}$  and  $0 < \lambda \leq \lambda_n \leq \frac{\delta-\alpha[\alpha(1+\alpha)+\alpha\delta+\tau]}{\delta[1+\alpha(1+\alpha)+\alpha\delta+\tau]}$ ,  $\forall n \geq 1$ .

Then

$$\mathbf{T}_n \mathbf{x}_n - \mathbf{x}_n \rightarrow 0 \text{ P-a.s.}$$

In addition, assume that:

(v)  $\mathfrak{Q}(\mathbf{x}_n)_{n \in \mathbb{N}} \subset \mathbf{D}$  P-a.s. Then  $(\mathbf{x}_n)_{n \in \mathbb{N}}$  converges weakly P-a.s. to a  $\mathbf{D}$ -valued random variable  $\hat{\mathbf{x}}$ . Furthermore, if

(vi)  $\mathfrak{Y}(\mathbf{x}_n)_{n \in \mathbb{N}} \neq \emptyset$  P-a.s.,

then  $(\mathbf{x}_n)_{n \in \mathbb{N}}$  converges strongly P-a.s. to  $\hat{\mathbf{x}}$ .

*Proof.* We define the norm  $||| \cdot |||$  on  $\mathbf{H}$  by

$$|||\mathbf{x}|||^2 = \sum_{i=1}^m \frac{1}{p_i} \|x_i\|^2, \quad \forall \mathbf{x} \in \mathbf{H}. \quad (3.3)$$

We will use Theorem 2.1 in  $(\mathbf{H}, ||| \cdot |||)$ . For every  $n \in \mathbb{N}$ , set  $\mathbf{t}_n = (t_{i,n})_{1 \leq i \leq m}$  where  $t_{i,n} = w_{i,n} + \varepsilon_{i,n}(T_{i,n}\mathbf{w}_n - w_{i,n})$ ,  $\forall n \in \mathbb{N}$ . Then from (3.2), we can know that

$$(\forall n \in \mathbb{N}) \begin{cases} \mathbf{w}_n = \mathbf{x}_n + \alpha_n(\mathbf{x}_n - \mathbf{x}_{n-1}), \\ \mathbf{x}_{n+1} = \mathbf{w}_n + \lambda_n(\mathbf{t}_n - \mathbf{w}_n). \end{cases} \quad (3.4)$$

Since the operators  $(\mathbf{T}_n)_{n \in \mathbb{N}}$  are quasinonexpansive,  $\mathbf{D}$  is closed [20, Section 2]. Now let  $\mathbf{y} \in \mathbf{D}$  and set

$$q_{i,n} =: \mathbf{H} \times M \rightarrow \mathbb{R} : (\mathbf{w}, \mathbf{s}) \mapsto \|w_i - y_i + \varepsilon_i(T_{i,n}\mathbf{w} - w_i)\|^2, \forall n \in \mathbb{N}, \forall i \in \{1, \dots, m\}. \quad (3.5)$$

Note that, for every  $n \in \mathbb{N}$  and every  $i \in \{1, \dots, m\}$ , since  $T_{i,n}$  is measurable, so are the functions  $(q_{i,n}(\cdot, \mathbf{s}))_{\mathbf{s} \in M}$ . With the same proof of Theorem 3.2 in [2], we can obtain that

$$\begin{aligned} (\forall n \in \mathbb{N}) E(|||\mathbf{t}_n - \mathbf{y}|||^2 | \mathcal{X}_n) &= \sum_{i=1}^m \frac{1}{p_i} E(\|t_{i,n} - y_i\|^2 | \mathcal{X}_n) \\ &= \sum_{i=1}^m \frac{1}{p_i} \sum_{\mathbf{s} \in M} P[\varepsilon_n = \mathbf{s}] \|w_{i,n} - y_i + \varepsilon_i(T_{i,n}\mathbf{w}_n - w_{i,n})\|^2 \end{aligned}$$

$$\leq \|\mathbf{w}_n - \mathbf{y}\|^2. \quad (3.6)$$

It is obvious that for every  $n \in \mathbb{N}$ , conditions (i)-(iii) of Theorem 2.1 are satisfied. Therefore, we can derive from conclusion of (iii) in Theorem 2.1 that  $\sum_{n \in \mathbb{N}} \mathbb{E}(\|\mathbf{t}_n - \mathbf{y}\|^2 | \mathcal{X}_n) < +\infty$  P-a.s. This yields

$$\mathbb{E}(\|\mathbf{t}_n - \mathbf{y}\|^2 | \mathcal{X}_n) \rightarrow 0 \quad (3.7)$$

P-a.s.

On the other hand, we can know that

$$\begin{aligned} (\forall n \in \mathbb{N}) \mathbb{E}(\|\mathbf{t}_n - \mathbf{w}_n\|^2 | \mathcal{X}_n) &= \sum_{i=1}^m \frac{1}{p_i} \mathbb{E}(\|t_{i,n} - w_{i,n}\|^2 | \mathcal{X}_n) \\ &= \|\mathbf{T}_n \mathbf{w}_n - \mathbf{w}_n\|^2. \end{aligned} \quad (3.8)$$

From (3.7), we have  $\mathbf{T}_n \mathbf{w}_n - \mathbf{w}_n \rightarrow 0$  P-a.s. In addition, by the consequences of Theorem 2.1, we can obtain the weak and strong convergence.  $\square$

**Definition 3.3.** Let  $T : H \rightarrow H$  be nonexpansive and let  $\beta \in (0, 1)$ . Then  $T$  is averaged with constant  $\beta$ , or  $\beta$ -averaged, if there exists a nonexpansive operator  $R : H \rightarrow H$  such that  $T = (1 - \beta)Id + \beta R$ .

**Proposition 3.4.** ([21, Proposition 4.25]). Let  $T : H \rightarrow H$  be nonexpansive and let  $\beta \in (0, 1)$ . Then  $T$  is  $\beta$ -averaged, if and only if

$$\|Tx - Ty\|^2 \leq \|x - y\|^2 - \frac{1 - \beta}{\beta} \|(I - T)x - (I - T)y\|^2, \quad \forall x, y \in H.$$

**Theorem 3.5.** Let  $(\forall n \in \mathbb{N}) \mathbf{T}_n : \mathbf{H} \rightarrow \mathbf{H} : \mathbf{x} \mapsto (T_{i,n}\mathbf{x})_{1 \leq i \leq m}$  be a  $\beta_n$ -averaged operator where, for every  $i \in \{1, \dots, m\}$ ,  $T_{i,n} : \mathbf{H} \rightarrow H_i$  and  $\beta_n$  is a sequence in  $(0, 1)$ . Suppose that  $b \in (0, 1)$ , and set  $M = \{0, 1\}^m \setminus \{\mathbf{0}\}$ . Let  $\mathbf{x}_0, \mathbf{x}_1$  be  $\mathbf{H}$ -valued random variables which are arbitrarily chosen, and let  $(\varepsilon_n)_{n \in \mathbb{N}}$  be identically distributed  $M$ -valued random variables. For  $n \geq 0$ ,

$$\begin{cases} \mathbf{w}_n = \mathbf{x}_n + \alpha_n(\mathbf{x}_n - \mathbf{x}_{n-1}), \\ x_{i,n+1} = w_{i,n} + \varepsilon_{i,n} \lambda_n(T_{i,n}(w_{1,n}, \dots, w_{m,n}) - w_{i,n}), i = \{1, \dots, m\}, \end{cases} \quad (3.9)$$

and set  $\mathcal{E}_n = \sigma(\varepsilon_n)$ . Moreover, suppose that there exists  $\hat{\Omega} \in \mathcal{F}$  such that  $P(\hat{\Omega}) = 1$  and the following hold:

- (i)  $\mathbf{D} = \bigcap_{n \in \mathbb{N}} \text{Fix}(\mathbf{T}_n) \neq \emptyset$ .
- (ii) For every  $n \in \mathbb{N}$ ,  $\mathcal{E}_n$  and  $\mathcal{X}_n$  are independent.
- (iii)  $P[\varepsilon_{i,0} = 1] > 0, \forall i \in \{1, \dots, m\}$ .
- (iv)  $(\alpha_n)_{n \geq 1}$  is nondecreasing with  $\alpha_1 = 0$  and  $0 \leq \alpha_n \leq \alpha < 1, \forall n \geq 1$  and  $\lambda, \tau, \delta > 0$  are such that  $\delta > \frac{\alpha^2(1+\alpha)+\alpha\tau}{1-\alpha^2}$  and  $0 < \max\{\lambda, \frac{b}{\beta_n}\} \leq \lambda_n \leq \min\{\frac{\delta-\alpha[\alpha(1+\alpha)+\alpha\delta+\tau]}{\delta[1+\alpha(1+\alpha)+\alpha\delta+\tau]}, \frac{(1-b)}{\beta_n}\}, \forall n \geq 1$ .
- (v)  $(\forall \omega \in \hat{\Omega}) [\beta_n^{-1}(\mathbf{T}_n \mathbf{x}_n(\omega) - \mathbf{x}_n(\omega)) \rightarrow 0 \implies \mathfrak{Q}(\mathbf{x}_n)_{n \in \mathbb{N}} \subset \mathbf{D}]$ . Then  $(\mathbf{x}_n)_{n \in \mathbb{N}}$  converges weakly  $P$ -a.s. to a  $\mathbf{D}$ -valued random variable  $\hat{\mathbf{x}}$ . Furthermore, if
- (vi)  $(\forall \omega \in \hat{\Omega}) [\sup_{n \in \mathbb{N}} \|\mathbf{x}_n(\omega)\| < +\infty \text{ and } \beta_n^{-1}(\mathbf{T}_n \mathbf{x}_n(\omega) - \mathbf{x}_n(\omega)) \rightarrow 0] \implies \mathfrak{Y}(\mathbf{x}_n)_{n \in \mathbb{N}} \neq \emptyset$ , then  $(\mathbf{x}_n)_{n \in \mathbb{N}}$  converges strongly  $P$ -a.s. to  $\hat{\mathbf{x}}$ .

*Proof.* Set  $\mathbf{V}_n = (1 - \beta_n^{-1})I + \beta_n^{-1}\mathbf{T}_n$  and  $\mathbf{V}_{i,n} = (1 - \beta_n^{-1})I + \beta_n^{-1}\mathbf{T}_{i,n}, \forall i \in \{1, \dots, m\}$ . Furthermore, for every  $n \in \mathbb{N}$ , we set  $\vartheta_n = \beta_n \lambda_n$ . So,  $\text{Fix}(\mathbf{V}_n) = \text{Fix}(\mathbf{T}_n), \forall n \in \mathbb{N}$  and  $\mathbf{V}_n$  is nonexpansive. Hence, for  $n \in \mathbb{N}$ , from (3.9) we have

$$\begin{cases} \mathbf{w}_n = \mathbf{x}_n + \alpha_n(\mathbf{x}_n - \mathbf{x}_{n-1}), \\ x_{i,n+1} = w_{i,n} + \varepsilon_{i,n}\vartheta_n(V_{i,n}(\mathbf{w}_n) - w_{i,n}), i = \{1, \dots, m\}. \end{cases} \quad (3.10)$$

Therefore, from Remark 3.3(iii) in [2] and Theorem 3.2, we can obtain the result.  $\square$

*Remark 3.1.* The binary variable  $\varepsilon_{i,n}$  in Theorem 3.2 and Theorem 3.5 signals whether the  $i$ -th coordinate  $T_{i,n}$  of the operator  $\mathbf{T}_n$  is activated or not at iteration  $n$ .

**Corollary 3.6.** Let  $(\beta_n)_{n \in \mathbb{N}}$  and  $(\gamma_n)_{n \in \mathbb{N}}$  be sequences in  $(0, 1)$  such that  $\sup_{n \in \mathbb{N}} \beta_n < 1$  and  $\sup_{n \in \mathbb{N}} \gamma_n < 1$ . Suppose that  $b \in (0, 1)$ , and set  $M = \{0, 1\}^m \setminus \{\mathbf{0}\}$ . Let  $\mathbf{x}_0, \mathbf{x}_1$  be  $\mathbf{H}$ -valued random variables which are arbitrarily chosen, and let  $(\varepsilon_n)_{n \in \mathbb{N}}$  be identically distributed  $M$ -valued random variables.  $\forall n \in \mathbb{N}$ , let  $\mathbf{V}_n : \mathbf{H} \rightarrow \mathbf{H}$  be a  $\gamma_n$ -averaged and  $\mathbf{T}_n : \mathbf{H} \rightarrow \mathbf{H} : \mathbf{x} \mapsto (T_{i,n}\mathbf{x})_{1 \leq i \leq m}$  be a  $\beta_n$ -averaged operator where,  $\forall i \in \{1, \dots, m\}$ ,  $T_{i,n} : \mathbf{H} \rightarrow H_i$ . For  $n \geq 0$ ,

$$\begin{cases} \mathbf{w}_n = \mathbf{x}_n + \alpha_n(\mathbf{x}_n - \mathbf{x}_{n-1}), \\ \mathbf{z}_n = \mathbf{V}_n \mathbf{w}_n, \\ x_{i,n+1} = w_{i,n} + \varepsilon_{i,n}\lambda_n(T_{i,n}(\mathbf{z}_n) - w_{i,n}), i = \{1, \dots, m\}, \end{cases} \quad (3.11)$$

- (i)  $\mathbf{D} = \bigcap_{n \in \mathbb{N}} \text{Fix}(\mathbf{T}_n \circ \mathbf{V}_n) \neq \emptyset$ .
- (ii) For every  $n \in \mathbb{N}$ ,  $\mathcal{E}_n$  and  $\mathcal{X}_n$  are independent.
- (iii)  $P[\varepsilon_{i,0} = 1] > 0, \forall i \in \{1, \dots, m\}$ .
- (iv)  $(\alpha_n)_{n \geq 1}$  is nondecreasing with  $\alpha_1 = 0$  and  $0 \leq \alpha_n \leq \alpha < 1, \forall n \geq 1$  and  $\lambda, \tau, \delta > 0$  are such that  $\delta > \frac{\alpha^2(1+\alpha)+\alpha\tau}{1-\alpha^2}$  and  $0 < \max\{\lambda, \frac{b}{\beta_n}\} \leq \lambda_n \leq \min\{\frac{\delta-\alpha[\alpha(1+\alpha)+\alpha\delta+\tau]}{\delta[1+\alpha(1+\alpha)+\alpha\delta+\tau]}, \frac{(1-b)}{\beta_n}\}, \forall n \geq 1$ .
- (v) Assuming that  $\mathfrak{Q}(\mathbf{x}_n)_{n \in \mathbb{N}} \subset \mathbf{D}$  *P*-a.s. Then  $(\mathbf{x}_n)_{n \in \mathbb{N}}$  converges weakly *P*-a.s. to a  $\mathbf{D}$ -valued random variable  $\hat{\mathbf{x}}$ .

*Proof.* Since  $\mathbf{V}_n$  is a  $\gamma_n$ -averaged operator and  $\mathbf{T}_n$  is a  $\beta_n$ -averaged operator, from Lemma 2.4, we can know that  $\forall n \in \mathbb{N}, \mathbf{S}_n = \mathbf{T}_n \circ \mathbf{V}_n$  is a  $\eta_n = \frac{\beta_n + \gamma_n - 2\beta_n\gamma_n}{1 - \beta_n\gamma_n}$ -averaged operator and  $\mathbf{S}_{i,n} = \mathbf{T}_{i,n} \circ \mathbf{V}_{i,n}, \forall i \in \{1, \dots, m\}$ . So,  $\text{Fix}(\mathbf{V}_n) = \text{Fix}(\mathbf{T}_n)$ , set  $\varpi_n = \eta_n \lambda_n$ . Hence, for  $n \in \mathbb{N}$ , from (3.11) we have

$$\begin{cases} \mathbf{w}_n = \mathbf{x}_n + \alpha_n(\mathbf{x}_n - \mathbf{x}_{n-1}), \\ x_{i,n+1} = w_{i,n} + \varepsilon_{i,n} \varpi_n (S_{i,n}(\mathbf{w}_n) - w_{i,n}), i = \{1, \dots, m\}. \end{cases} \quad (3.12)$$

Therefore, we can obtain the result from Theorem 3.5. □

## 4 A preconditioned stochastic inertial block-coordinate forward-backward algorithm

First, we introduce some definitions and notations. We refer the readers to [21] for more details. Let  $\tilde{\mathbf{M}} : \mathbf{H} \rightarrow \mathbf{H}$  be a set-valued operator. We denote by  $\text{ran}(\tilde{\mathbf{M}}) := \{v \in \mathbf{H} : \exists u \in \mathbf{H}, v \in \tilde{\mathbf{M}}u\}$  the range of  $\tilde{\mathbf{M}}$ , by  $\text{gra}(\tilde{\mathbf{M}}) := \{(u, v) \in \mathbf{H}^2 : v \in \tilde{\mathbf{M}}u\}$  its graph, and by  $\tilde{\mathbf{M}}^{-1}$  its inverse; that is, the set-valued operator with graph  $\{(v, u) \in \mathbf{H}^2 : v \in \tilde{\mathbf{M}}u\}$ . We define  $\text{zer}(\tilde{\mathbf{M}}) := \{u \in \mathbf{H} : 0 \in \tilde{\mathbf{M}}u\}$ .  $\tilde{\mathbf{M}}$  is said to be monotone if  $\forall (u, u') \in \mathbf{H}^2, \forall (v, v') \in \tilde{\mathbf{M}}u \times \tilde{\mathbf{M}}u', \langle u - u', v - v' \rangle \geq 0$  and maximally monotone if there exists no monotone operator  $\tilde{\mathbf{M}}'$  such that  $\text{gra}(\tilde{\mathbf{M}}) \subset \text{gra}(\tilde{\mathbf{M}}') \neq \text{gra}(\tilde{\mathbf{M}})$ .

The resolvent  $(\mathbf{I} + \tilde{\mathbf{M}})^{-1}$  of a maximally monotone operator  $\tilde{\mathbf{M}} : \mathbf{H} \rightarrow \mathbf{H}$  is defined and single-valued on  $\mathbf{H}$  and firmly nonexpansive. The subdifferential  $\partial \mathbf{J}$  of  $\mathbf{J} \in \Gamma_0(\mathbf{H})$  is maximally monotone.

**Theorem 4.1.** Let  $\mathbf{A} : \mathbf{H} \rightarrow 2^{\mathbf{H}}$  be a maximally monotone operator and let  $\mathbf{B} : \mathbf{H} \rightarrow 2^{\mathbf{H}}$  be a cocoercive operator. Assume that  $\mathbf{D} = \text{zer}(\mathbf{A} + \mathbf{B})$  is nonempty. Let  $\tilde{\mathbf{L}}$  be a strongly positive self-adjoint operator in  $\mathcal{B}(\mathbf{H})$  such that  $\tilde{\mathbf{L}}^{1/2}\mathbf{B}\tilde{\mathbf{L}}^{1/2}$  is  $\mu$ -cocoercive with  $\mu \in (0, +\infty)$ . Let  $(\theta_n)_{n \in \mathbb{N}}$  be a sequence in  $\mathbb{R}$  such that  $0 < \inf_{n \in \mathbb{N}} \theta_n \leq \sup_{n \in \mathbb{N}} \theta_n < 2\mu$  and set  $M = \{0, 1\}^m \setminus \{\mathbf{0}\}$ . Let  $\mathbf{x}_0, \mathbf{x}_1$  be  $\mathbf{H}$ -valued random variables which are arbitrarily chosen, and let  $(\varepsilon_n)_{n \in \mathbb{N}}$  be identically distributed  $M$ -valued random variables.  $\forall n \in \mathbb{N}$ , set  $\mathbf{J}_{\theta_n \tilde{\mathbf{L}} \mathbf{A}} : \mathbf{x} \mapsto (T_{i,n} \mathbf{x})_{1 \leq i \leq m}$ , where  $\forall i \in \{1, \dots, m\}$ ,  $T_{i,n} : \mathbf{H} \rightarrow \mathbf{H}_i$ . For  $n \geq 0$ :

$$\begin{cases} \mathbf{w}_n = \mathbf{x}_n + \alpha_n(\mathbf{x}_n - \mathbf{x}_{n-1}), \\ \mathbf{z}_n = \tilde{\mathbf{L}} \mathbf{A} \mathbf{w}_n, \\ x_{i,n+1} = w_{i,n} + \varepsilon_{i,n} \lambda_n (T_{i,n}(\mathbf{w}_n - \theta_n \mathbf{z}_n) - w_{i,n}), i = \{1, \dots, m\}. \end{cases} \quad (4.1)$$

(i) For every  $n \in \mathbb{N}$ ,  $\mathcal{E}_n$  and  $\mathcal{X}_n$  are independent.

(ii)  $P[\varepsilon_{i,0} = 1] > 0$ ,  $\forall i \in \{1, \dots, m\}$ .

(iii)  $(\alpha_n)_{n \geq 1}$  is nondecreasing with  $\alpha_1 = 0$  and  $0 \leq \alpha_n \leq \alpha < 1$ ,  $\forall n \geq 1$  and  $\lambda, \tau, \delta > 0$  are such that  $\delta > \frac{\alpha^2(1+\alpha)+\alpha\tau}{1-\alpha^2}$  and  $0 < \lambda \leq \lambda_n \leq \frac{\delta - \alpha[\alpha(1+\alpha)+\alpha\delta+\tau]}{\delta[1+\alpha(1+\alpha)+\alpha\delta+\tau]}$ ,  $\forall n \geq 1$ .

Then  $(\mathbf{x}_n)_{n \in \mathbb{N}}$  converges weakly  $P$ -a.s. to a  $\mathbf{D}$ -valued random variable  $\hat{\mathbf{x}}$ .

*Proof.* By assumption, we can know that  $\text{zer}(\tilde{\mathbf{L}} \mathbf{A} + \tilde{\mathbf{L}} \mathbf{B}) = \text{zer}(\mathbf{A} + \mathbf{B}) \neq \emptyset$ . Since  $\tilde{\mathbf{L}}$  is a strongly positive self-adjoint operator, we can define a particular inner product  $\langle \cdot, \cdot \rangle_{\tilde{\mathbf{L}}^{-1}}$  and norm  $\| \cdot \|_{\tilde{\mathbf{L}}^{-1}} = \langle \cdot, \cdot \rangle_{\tilde{\mathbf{L}}^{-1}}^{\frac{1}{2}}$  in  $\mathbf{H}$  as

$$\langle \mathbf{x}, \mathbf{x}' \rangle_{\tilde{\mathbf{L}}^{-1}} = \langle \mathbf{x}, \tilde{\mathbf{L}}^{-1} \mathbf{x}' \rangle, \forall \mathbf{x}, \mathbf{x}' \in \mathbf{H}. \quad (4.2)$$

By endowing  $\mathbf{H}$  with this inner product, we obtain the Hilbert space denoted by  $\mathbf{H}_{\tilde{\mathbf{L}}^{-1}}$ . In this renormed space,  $\tilde{\mathbf{L}} \mathbf{A}$  is maximally monotone. In addition, for every  $\mathbf{x}, \mathbf{x}' \in \mathbf{H}$ , from the proof of Proposition 3.1 in [3], we have

$$\begin{aligned} \|\tilde{\mathbf{L}} \mathbf{B} \mathbf{x} - \tilde{\mathbf{L}} \mathbf{B} \mathbf{x}'\|_{\tilde{\mathbf{L}}^{-1}}^2 &= \|\tilde{\mathbf{L}}^{\frac{1}{2}} \mathbf{B} \mathbf{x} - \tilde{\mathbf{L}}^{\frac{1}{2}} \mathbf{B} \mathbf{x}'\|^2 \\ &\leq \mu^{-1} \langle \mathbf{x} - \mathbf{x}', \tilde{\mathbf{L}} \mathbf{B} \mathbf{x} - \tilde{\mathbf{L}} \mathbf{B} \mathbf{x}' \rangle_{\tilde{\mathbf{L}}^{-1}}, \end{aligned} \quad (4.3)$$

which shows that  $\tilde{\mathbf{L}} \mathbf{B}$  is  $\mu$ -cocoercive in  $(\mathbf{H}, \| \cdot \|_{\tilde{\mathbf{L}}^{-1}})$ . Hence, we can find an element of  $\mathbf{Z}$  by composing operators  $\mathbf{J}_{\theta_n \tilde{\mathbf{L}} \mathbf{A}}$  and  $\mathbf{I} - \theta_n \tilde{\mathbf{L}} \mathbf{B}$ . Since the first operator is  $1/2$ -averaged and the second one is  $\theta_n/(2\mu)$ -averaged [21, Proposition 4.33]. Noticing that weak convergences in the sense of  $\langle \cdot, \cdot \rangle$  and  $\langle \cdot, \cdot \rangle_{\tilde{\mathbf{L}}^{-1}}$  are equivalent. So, from the Corollary 3.6, we can obtain the convergence result.  $\square$

## 5 Applications

### 5.1 Intertial block-coordinate primal-dual algorithms for monotone inclusion problems

Let  $(E_j)_{1 \leq j \leq p}$  and  $(G_k)_{1 \leq k \leq q}$  be separable real Hilbert spaces, where  $p, q$  are positive integers. Furthermore,  $\mathbf{E} = E_1 \oplus \cdots \oplus E_p$  and  $\mathbf{G} = G_1 \oplus \cdots \oplus G_q$  denote the Hilbert direct sums of  $(E_j)_{1 \leq j \leq p}$  and  $(G_k)_{1 \leq k \leq q}$ , respectively. We take into account  $\mathbf{H} = \mathbf{E} \oplus \mathbf{G}$ . Recently, more and more people pay much attention to the problem involving monotone operators (see e.g. [4-9]), it also play a significant role in our work.

**Problem 5.1.** *Let  $A_j : E_j \rightarrow 2^{E_j}$  be maximally monotone, and  $C_j : E_j \rightarrow E_j$  be cocoercive,  $\forall j \in \{1, \dots, p\}$ . For every  $k \in \{1, \dots, q\}$ , let  $B_k : G_k \rightarrow 2^{G_k}$  be maximally monotone, let  $\tilde{D}_k : G_k \rightarrow 2^{G_k}$  be maximally and strongly monotone, and let  $L_{k,j} \in \mathcal{B}(E_j, G_k)$ . Assuming that*

$$\mathbb{L}_k = \{j \in \{1, \dots, p\} | L_{k,j} \neq 0\} \neq \emptyset, \forall k \in \{1, \dots, q\}, \quad (5.1)$$

$$\mathbb{L}_j^* = \{k \in \{1, \dots, q\} | L_{k,j} \neq 0\} \neq \emptyset, \forall j \in \{1, \dots, p\}, \quad (5.2)$$

and that the set  $\mathbf{Z}$  of solutions to the problem:

find  $x_1 \in E_1, \dots, x_p \in E_p$  such that

$$0 \in A_j x_j + C_j x_j + \sum_{k=1}^q L_{k,j}^* (B_k \square \tilde{D}_k) \left( \sum_{j'=1}^p L_{k,j'} x_{j'} \right) \quad (5.3)$$

is nonempty. Furthermore, we consider the set  $\mathbf{Z}^*$  of solutions to the dual problem:

find  $y_1 \in G_1, \dots, y_q \in G_q$  such that

$$0 \in - \sum_{j=1}^p L_{k,j} (A_j^{-1} \square C_j^{-1}) \left( - \sum_{k'=1}^q L_{k',j}^* y_{k'} \right) + B_k^{-1} y_k + \tilde{D}_k^{-1} y_k. \quad (5.4)$$

We aim at finding a pair  $(\hat{\mathbf{x}}, \hat{\mathbf{y}})$  of random variables such that  $\hat{\mathbf{x}}$  is  $\mathbf{Z}$ -valued and  $\hat{\mathbf{y}}$  is  $\mathbf{Z}^*$ -valued.

By [3, 10-11], we can know that the above problem can be regard as a search for a zero of the sum of two maximally monotone operators in the product space  $\mathbf{H}$ .

**Lemma 5.1.** ([3]). Let  $\mathbf{A} : \mathbf{E} \rightarrow 2^{\mathbf{E}} : \mathbf{x} \mapsto \times_{j=1}^p A_j x_j$ ,  $\mathbf{B} : \mathbf{G} \rightarrow 2^{\mathbf{G}} : \mathbf{y} \mapsto \times_{k=1}^q B_k y_k$ ,  $\mathbf{C} : \mathbf{E} \rightarrow 2^{\mathbf{E}} : \mathbf{x} \mapsto (C_j x_j)_{1 \leq j \leq p}$ ,  $\tilde{\mathbf{D}} : \mathbf{G} \rightarrow 2^{\mathbf{G}} : \mathbf{y} \mapsto \times_{k=1}^q \tilde{D}_k y_k$ , and  $\mathbf{L} : \mathbf{E} \rightarrow \mathbf{G} : \mathbf{x} \mapsto (\sum_{j=1}^p L_{k,j} x_j)_{1 \leq k \leq q}$ . Now, we consider the operators

$$\mathbf{U} : \begin{pmatrix} \mathbf{x} \\ \mathbf{y} \end{pmatrix} \mapsto \begin{pmatrix} \mathbf{A} & \mathbf{L}^* \\ -\mathbf{L} & \mathbf{B}^{-1} \end{pmatrix} \begin{pmatrix} \mathbf{x} \\ \mathbf{y} \end{pmatrix}, \quad (5.5)$$

and

$$\mathbf{V} : \begin{pmatrix} \mathbf{x} \\ \mathbf{y} \end{pmatrix} \mapsto \begin{pmatrix} \mathbf{C} \\ \tilde{\mathbf{D}}^{-1} \end{pmatrix} \begin{pmatrix} \mathbf{x} \\ \mathbf{y} \end{pmatrix}. \quad (5.6)$$

Then, the following hold:

- (i)  $\mathbf{U}$  is maximally monotone and  $\mathbf{V}$  is cocoercive.
- (ii)  $\mathbf{D} = \text{zer}(\mathbf{U} + \mathbf{V})$  is nonempty.
- (iii) A pair  $(\hat{\mathbf{x}}, \hat{\mathbf{y}})$  of random variables is a solution to Problem 5.1 if and only if  $(\hat{\mathbf{x}}, \hat{\mathbf{y}})$  is  $\mathbf{D}$ -valued.

Now, we will consider an inertial block-coordinate primal-dual algorithms for Problem 5.1.

**Theorem 5.1.** Let  $\mathbf{F} : \mathbf{E} \rightarrow \mathbf{E} : \mapsto (F_1 x_1, \dots, F_p x_p)$  and  $\mathbf{R} : \mathbf{G} \rightarrow \mathbf{G} : \mapsto (R_1 y_1, \dots, R_q y_q)$  where,  $\forall j \in \{1, \dots, p\}$ ,  $F_j$  is a strongly positive self-adjoint operator in  $\mathcal{B}(E_j)$  such that  $F_j^{\frac{1}{2}} C_j F_j^{\frac{1}{2}}$  is  $\nu_j$ -cocoercive with  $\nu_j \in (0, +\infty)$ , and  $\forall k \in \{1, \dots, q\}$ ,  $R_k$  is a strongly positive self-adjoint operator in  $\mathcal{B}(G_k)$  such that  $R_k^{\frac{1}{2}} \tilde{D}_k R_k^{\frac{1}{2}}$  is  $\tilde{\tau}_k$ -cocoercive with  $\tilde{\tau}_k \in (0, +\infty)$ . Assume that  $(\exists a \in (0, +\infty)) 2\mu_a > 1$  where the definition of  $\mu_a$  is similar with the definition of  $\vartheta_a$  in Lemma 4.3 [3] with  $\nu = \min\{\nu_1, \dots, \nu_p\}$  and  $\tilde{\tau} = \min\{\tilde{\tau}_1, \dots, \tilde{\tau}_q\}$ . Let  $\mathbf{x}_0, \mathbf{x}_1$  be  $\mathbf{E}$ -valued random variables which are arbitrarily chosen, and let  $\mathbf{y}_0, \mathbf{y}_1$  be  $\mathbf{G}$ -valued random variables which are arbitrarily chosen. Let  $(\varepsilon_n)_{n \in \mathbb{N}}$  be identically dis-



tributed  $M_{p+q}$ -valued random variables. For  $n \geq 0$ :

$$\left\{ \begin{array}{l} \text{for } j = \dots, p, \\ w_{j,n} = x_{j,n} + \alpha_n(x_{j,n} - x_{j,n-1}), \\ z_{j,n} = \varepsilon_{j,n}(J_{F_j A_j}(w_{j,n} - F_j(\sum_{k \in \mathbb{L}_j^*} L_{k,j}^* y_{k,n} + C_j w_{j,n}))), \\ x_{j,n+1} = w_{j,n} + \varepsilon_{j,n} \lambda_n(z_{j,n} - w_{j,n}), \\ \text{for } k = \dots, q, \\ h_{k,n} = y_{k,n} + \alpha_n(y_{k,n} - y_{k,n-1}), \\ s_{k,n} = \varepsilon_{p+k,n}(J_{R_k B_k^{-1}}(h_{k,n} + R_k(\sum_{j \in \mathbb{L}_k} L_{k,j}(2z_{j,n} - w_{j,n}) - \tilde{D}_k^{-1} h_{k,n}))), \\ y_{k,n+1} = h_{k,n} + \varepsilon_{p+k,n} \lambda_n(s_{k,n} - h_{k,n}), \end{array} \right. \quad (5.7)$$

and set  $\mathcal{E}_n = \sigma(\varepsilon_n)$ ,  $\tilde{\mathcal{X}}_n = \sigma(\mathbf{x}_{n'}, \mathbf{y}_{n'})_{0 \leq n' \leq n}$ . Moreover, suppose that the following hold:

(i) For every  $n \in \mathbb{N}$ ,  $\mathcal{E}_n$  and  $\tilde{\mathcal{X}}_n$  are independent, and  $P[\varepsilon_{p+k,0} = 1] > 0$ ,  $\forall k \in \{1, \dots, q\}$ .  
(ii) For every  $j \in \{1, \dots, p\}$  and  $n \in \mathbb{N}$ ,  $\bigcup_{k \in \mathbb{L}_j^*} \{\omega \in \Omega | \varepsilon_{p+k,n}(\omega) = 1\} \subset \{\omega \in \Omega | \varepsilon_{j,n}(\omega) = 1\}$ .

(iii)  $(\alpha_n)_{n \geq 1}$  is nondecreasing with  $\alpha_1 = 0$  and  $0 \leq \alpha_n \leq \alpha < 1$ ,  $\forall n \geq 1$  and  $\lambda, \tau, \delta > 0$  are such that  $\delta > \frac{\alpha^2(1+\alpha)+\alpha\tau}{1-\alpha^2}$  and  $0 < \lambda \leq \lambda_n \leq \frac{\delta-\alpha[\alpha(1+\alpha)+\alpha\delta+\tau]}{\delta[1+\alpha(1+\alpha)+\alpha\delta+\tau]}$ ,  $\forall n \geq 1$ .

Then  $(\mathbf{x}_n)_{n \in \mathbb{N}}$  converges weakly  $P$ -a.s. to a  $\mathbf{Z}$ -valued random variable  $\hat{\mathbf{x}}$ , and  $(\mathbf{y}_n)_{n \in \mathbb{N}}$  converges weakly  $P$ -a.s. to a  $\mathbf{Z}^*$ -valued random variable  $\hat{\mathbf{y}}$ .

*Proof.* By Lemma 5.1(i)-(ii), we can know that  $\mathbf{U}$  is maximally monotone,  $\mathbf{V}$  is coco-ercive, and  $\mathbf{D} = \text{zer}(\mathbf{U} + \mathbf{V}) \neq \emptyset$ . On the other hand,  $(\exists a \in (0, +\infty)) 2\mu_a > 1$  and the definition of  $\mu_a$  imply that  $\|\mathbf{F}^{\frac{1}{2}} \mathbf{L} \mathbf{R}^{\frac{1}{2}}\| < 1$ . So, with the same idea of Lemma 4.5[3], Algorithm (5.7) can be rewritten under the form of Algorithm (4.1), where  $m = p + q$ ,  $\tilde{\mathbf{L}}$  is defined by (5.8)

$$\tilde{\mathbf{L}} : \begin{pmatrix} \mathbf{x} \\ \mathbf{y} \end{pmatrix} \mapsto \begin{pmatrix} (\mathbf{F}^{-1} - \mathbf{L}^* \mathbf{R} \mathbf{L})^{-1} & \mathbf{F} \mathbf{L}^* (\mathbf{R}^{-1} - \mathbf{L} \mathbf{F} \mathbf{L}^*)^{-1} \\ (\mathbf{R}^{-1} - \mathbf{L} \mathbf{F} \mathbf{L}^*)^{-1} \mathbf{L} \mathbf{F} & (\mathbf{R}^{-1} - \mathbf{L} \mathbf{F} \mathbf{L}^*)^{-1} \end{pmatrix} \begin{pmatrix} \mathbf{x} \\ \mathbf{y} \end{pmatrix}, \quad (5.8)$$

the more detail about  $\tilde{\mathbf{L}}$  can see ([3, Lemma 4.3]), and for every  $n \in \mathbb{N}$

$$\mathbf{x}_n = (\mathbf{x}_n, \mathbf{y}_n), \quad (5.9)$$

$$\theta_n = 1, \quad (5.10)$$

$$\mathbf{J}_{\tilde{\mathbf{L}}\mathbf{A}} : \mathbf{x} \mapsto (T_{i,n}\mathbf{x})_{1 \leq i \leq m}, \quad (5.11)$$

$$T_{j,n} : \mathbf{H} \rightarrow E_j, \forall j \in \{1, \dots, p\}, \quad (5.12)$$

$$T_{p+k,n} : \mathbf{H} \rightarrow G_k, \forall k \in \{1, \dots, q\}. \quad (5.13)$$

From Lemma 4.3(i) in [3] we can know that  $\tilde{\mathbf{L}}$  is a strongly positive self-adjoint operator in  $\mathcal{B}(\mathbf{H})$ . Therefore, with the same proof of Proposition 4.6 [3], we can know all the assumptions of Theorem 4.1 are satisfied, which allows us to establish the almost sure convergence of  $(\mathbf{x}_n, \mathbf{y}_n)$  to a  $\mathbf{D}$ -valued random variable. Finally, Lemma 5.1(iii) ensures that the limit is an  $\mathbf{Z} \times \mathbf{Z}^*$ -valued random variable.  $\square$

## 5.2 Inertial block-coordinate primal-dual algorithms for convex optimization problems

In this section, we will introduce an inertial block-coordinate primal-dual algorithms for solving a wide range of structured convex optimization problems. The results obtained in the previous section. In particular, we will pay our attention to the following optimization problems. We still use the notation of the previous section and present some new notations. We denote by  $\Gamma_0(H)$  the class of lower semicontinuous convex functions  $f : H \rightarrow (-\infty, +\infty)$  such that  $f \neq +\infty$ . The Moreau subdifferential of  $f \in \Gamma_0(H)$  is the maximally monotone operator

$$\partial f : H \rightarrow 2^H : x \mapsto \{u \in H \mid \langle y - x, u \rangle + f(x) \leq f(y)\}. \quad (5.14)$$

**Definition 5.2.** *Let  $f$  be a real-valued convex function on  $H$ , the operator  $\text{prox}_f$  is defined by*

$$\begin{aligned} \text{prox}_f : H &\rightarrow H \\ x &\mapsto \arg \min_{y \in \mathcal{X}} f(y) + \frac{1}{2} \|x - y\|_2^2, \end{aligned}$$

*called the proximity operator of  $f$ .*

For more details about convex analysis and monotone operator theory, see[21].

**Problem 5.2.**  $\forall j \in \{1, \dots, p\}$ , let  $f_j, h_j \in \Gamma_0(E_j)$ , and  $h_j$  be Lipschitz-differentiable.  $\forall k \in \{1, \dots, q\}$ , let  $g_k, l_k \in \Gamma_0(G_k)$ , and  $l_k$  be strongly convex. Let  $L_{k,j} \in \mathcal{B}(E_j, G_k)$ . Assume that (5.1) and (5.2) hold, and that there exists  $(\bar{x}_1, \dots, \bar{x}_p) \in E_1 \oplus \dots \oplus E_p$  such that

$$0 \in \partial f_j(\bar{x}_j) + \nabla h_j(\bar{x}_j) + \sum_{k=1}^q L_{k,j}^*(\partial g_k \square \partial l_k) \left( \sum_{j'=1}^p L_{k,j'} \bar{x}_{j'} \right), \quad \forall j \in \{1, \dots, p\}. \quad (5.15)$$

Let  $\tilde{\mathbf{Z}}$  be the set of solutions to the problem

$$\min_{x_1 \in E_1, \dots, x_p \in E_p} \sum_{j=1}^p (f_j(x_j) + h_j(x_j)) + \sum_{k=1}^q (g_k \square l_k) \left( \sum_{j=1}^p L_{k,j} x_j \right), \quad (5.16)$$

and let  $\tilde{\mathbf{Z}}^*$  be the set of solutions to the dual problem

$$\min_{y_1 \in G_1, \dots, y_q \in G_q} \sum_{j=1}^p (f_j^* \square h_j^*) \left( - \sum_{k=1}^q L_{k,j}^* y_k \right) + \sum_{k=1}^q (g_k^*(y_k) + l_k^*(y_k)). \quad (5.17)$$

We aim at finding a pair  $(\hat{\mathbf{x}}, \hat{\mathbf{y}})$  of random variables such that  $\hat{\mathbf{x}}$  is  $\tilde{\mathbf{Z}}$ -valued and  $\hat{\mathbf{y}}$  is  $\tilde{\mathbf{Z}}^*$ -valued.

In order to satisfy the condition in Problem 5.2, we need the following assumptions:

**Proposition 5.3.** ([10, Proposition 5.3]). Consider the setting of Problem 5.2. Suppose that (5.16) has a solution. Then, the existence of  $(\bar{x}_1, \dots, \bar{x}_p) \in E_1 \oplus \dots \oplus E_p$  satisfying (5.15) is guaranteed in each of the following cases:

- (i)  $\forall j \in \{1, \dots, p\}$ ,  $f_j$  is real-valued and  $\forall k \in \{1, \dots, q\}$ ,  $(x_j)_{1 \leq j \leq p} \mapsto L_{k,j} x_j$  is surjective.
- (ii)  $\forall k \in \{1, \dots, q\}$ ,  $g_k$  or  $l_k$  is real-valued.
- (iii)  $(E_j)_{1 \leq j \leq p}$  and  $(G_k)_{1 \leq k \leq q}$  are finite-dimensional, and  $\exists x_j \in \text{ri dom } f_j$  such that  $L_{k,j} x_j \in \text{ri dom } g_k + \text{ri dom } l_k$ .

The following result can be deduced from Theorem 5.1:

**Theorem 5.4.** Let  $\mathbf{F}$  and  $\mathbf{R}$  be defined as in Theorem 5.1.  $\forall j \in \{1, \dots, p\}$ , let  $\nu_j^{-1} \in (0, +\infty)$  be a Lipschitz constant of the gradient of  $h_j \circ F_j^{\frac{1}{2}}$ , and  $\forall k \in \{1, \dots, q\}$ , let

$\tilde{\tau}_k^{-1} \in (0, +\infty)$  be a Lipschitz constant of the gradient of  $l_k^* \circ R_k^{\frac{1}{2}}$ . Assume that  $(\exists a \in (0, +\infty)) \ 2\mu_a > 1$  where the definition of  $\mu_a$  is similar with the definition of  $\vartheta_a$  in Lemma 4.3 [3] with  $\nu = \min\{\nu_1, \dots, \nu_p\}$  and  $\tilde{\tau} = \min\{\tilde{\tau}_1, \dots, \tilde{\tau}_q\}$ . Let  $\mathbf{x}_0, \mathbf{x}_1$  be  $\mathbf{E}$ -valued random variables which are arbitrarily chosen, and let  $\mathbf{y}_0, \mathbf{y}_1$  be  $\mathbf{G}$ -valued random variables which are arbitrarily chosen. Let  $(\varepsilon_n)_{n \in \mathbb{N}}$  be identically distributed  $M_{p+q}$ -valued random variables. For  $n \geq 0$ :

$$\left\{ \begin{array}{l} \text{for } j = \dots, p, \\ w_{j,n} = x_{j,n} + \alpha_n(x_{j,n} - x_{j,n-1}), \\ z_{j,n} = \varepsilon_{j,n}(\text{prox}_{f_j}^{F_j^{-1}}(w_{j,n} - F_j(\sum_{k \in \mathbb{L}_j^*} L_{k,j}^* y_{k,n} + \nabla h_j(w_{j,n})))), \\ x_{j,n+1} = w_{j,n} + \varepsilon_{j,n} \lambda_n(z_{j,n} - w_{j,n}), \\ \text{for } k = \dots, q, \\ \tilde{h}_{k,n} = y_{k,n} + \alpha_n(y_{k,n} - y_{k,n-1}), \\ s_{k,n} = \varepsilon_{p+k,n}(\text{prox}_{g_k^*}^{B_k^{-1}}(\tilde{h}_{k,n} + R_k(\sum_{j \in \mathbb{L}_k} L_{k,j}(2z_{j,n} - w_{j,n}) - \nabla l_k^*(\tilde{h}_{k,n})))), \\ y_{k,n+1} = \tilde{h}_{k,n} + \varepsilon_{p+k,n} \lambda_n(s_{k,n} - \tilde{h}_{k,n}), \end{array} \right. \quad (5.18)$$

and set  $\mathcal{E}_n = \sigma(\varepsilon_n)$ ,  $\tilde{\mathcal{X}}_n = \sigma(\mathbf{x}_{n'}, \mathbf{y}_{n'})_{0 \leq n' \leq n}$ . Moreover, suppose that Conditions (i)-(iii) in Theorem 5.1 hold.

Then  $(\mathbf{x}_n)_{n \in \mathbb{N}}$  converges weakly  $P$ -a.s. to a  $\tilde{\mathbf{Z}}$ -valued random variable  $\hat{\mathbf{x}}$ , and  $(\mathbf{y}_n)_{n \in \mathbb{N}}$  converges weakly  $P$ -a.s. to a  $\tilde{\mathbf{Z}}^*$ -valued random variable  $\hat{\mathbf{y}}$ .

*Proof.*  $\forall j \in \{1, \dots, p\}$ , we set  $A_j = \partial f_j$ ,  $C_j = \nabla h_j$ , and  $\forall k \in \{1, \dots, q\}$ ,  $B_k = \partial g_k$ ,  $D_k^{-1} = \nabla l_k^*$ . Observing that  $\forall j \in \{1, \dots, p\}$  and  $\forall k \in \{1, \dots, q\}$ ,  $J_{F_j A_j} = \text{prox}_{f_j}^{F_j^{-1}}$ ,  $J_{R_k B_k^{-1}} = \text{prox}_{g_k^*}^{B_k^{-1}}$ . In addition, the Lipschitz-differentiability assumptions made on  $h_j$  and  $l_k^*$  are equivalent to the fact that  $F_j^{\frac{1}{2}} C_j F_j^{\frac{1}{2}}$  is  $\nu_j$ -cocoercive and  $R_k^{\frac{1}{2}} \tilde{D}_k^{-1} R_k^{\frac{1}{2}}$  is  $\tilde{\tau}_k$ -cocoercive.[21, Corollaries 16.42, 18.16]. The convergence result follows from [1, Proposition 5.3].

□

In Problem 5.2, if  $\forall j \in \{1, \dots, p\}$ ,  $f_j = 0$ , we can obtain the following Corollary.

**Corollary 5.5.** *Let  $\mathbf{F}$  and  $\mathbf{R}$  be defined as in Theorem 5.1. Let  $\nu$  and  $\tilde{\tau}$  be defined as in Theorem 5.4. Suppose that Condition  $\min\{\nu, \tilde{\tau}(1 - \|\mathbf{R}^{\frac{1}{2}} \mathbf{L} \mathbf{F}^{\frac{1}{2}}\|^2)\} > \frac{1}{2}$  holds. Let  $\mathbf{x}_0, \mathbf{x}_1$*

be  $\mathbf{E}$ -valued random variables which are arbitrarily chosen, and let  $\mathbf{y}_0, \mathbf{y}_1$  be  $\mathbf{G}$ -valued random variables which are arbitrarily chosen. Let  $(\varepsilon_n)_{n \in \mathbb{N}}$  be identically distributed  $M_{p+q}$ -valued random variables. For  $n \geq 0$ :

$$\left\{ \begin{array}{l} \text{for } j = \dots, p, \\ \xi_{j,n} = \max\{\varepsilon_{p+k,n} | k \in \mathbb{L}_j^*\}, \\ w_{j,n} = x_{j,n} + \alpha_n(x_{j,n} - x_{j,n-1}), \\ z_{j,n} = \xi_{j,n}(w_{j,n} - F_j \nabla h_j(w_{j,n})), \\ \text{for } k = \dots, q, \\ \tilde{h}_{k,n} = y_{k,n} + \alpha_n(y_{k,n} - y_{k,n-1}), \\ s_{k,n} = \varepsilon_{p+k,n}(\text{prox}_{g_k^*}^{B_k^{-1}}(\tilde{h}_{k,n} + R_k(\sum_{j \in \mathbb{L}_k} L_{k,j}(z_{j,n} - F_j \sum_{k' \in \mathbb{L}_j^*} L_{k',j}^* y_{k',n}) - \nabla l_k^*(\tilde{h}_{k,n})))), \\ y_{k,n+1} = \tilde{h}_{k,n} + \varepsilon_{p+k,n} \lambda_n(s_{k,n} - \tilde{h}_{k,n}), \\ \text{for } j = \dots, p, \\ x_{j,n+1} = w_{j,n} + \varepsilon_{j,n} \lambda_n(z_{j,n} - F_j \sum_{k \in \mathbb{L}_j^*} L_{k,j}^* s_{k,n} - w_{j,n}), \end{array} \right. \quad (5.19)$$

and set  $\mathcal{E}_n = \sigma(\varepsilon_n)$ ,  $\tilde{\mathcal{X}}_n = \sigma(\mathbf{x}_{n'}, \mathbf{y}_{n'})_{0 \leq n' \leq n}$ . Moreover, suppose that Condition (i) and (iii) in Theorem 5.1 is satisfied and the following hold:

For every  $k \in \{1, \dots, q\}$  and  $n \in \mathbb{N}$ ,  $\bigcup_{j \in \mathbb{L}_k} \{\omega \in \Omega | \varepsilon_{j,n}(\omega) = 1\} \subset \{\omega \in \Omega | \varepsilon_{p+k,n}(\omega) = 1\}$ .

Then  $(\mathbf{x}_n)_{n \in \mathbb{N}}$  converges weakly  $P$ -a.s. to a  $\tilde{\mathbf{Z}}$ -valued random variable  $\hat{\mathbf{x}}$ , and  $(\mathbf{y}_n)_{n \in \mathbb{N}}$  converges weakly  $P$ -a.s. to a  $\tilde{\mathbf{Z}}^*$ -valued random variable  $\hat{\mathbf{y}}$ .

*Remark 5.1.* Our results improve and extend the results of other people in the following aspects.

(i) If  $\alpha_n = 0, \forall n \in \mathbb{N}$  in Theorem 5.1, Theorem 5.4 and Corollary 5.5, we can obtain the Proposition 4.6, Proposition 5.3 and Proposition 5.4 of Jean-Christophe and Audrey [3] in the absence of errors.

(ii) If  $\forall n \in \mathbb{N}, \alpha_n = 0$  and  $p = 1$ , Algorithm (5.18) extends the deterministic approaches in [11-15] by introducing some random sweeping of the coordinates in the absence of errors. Similarly, when  $\alpha_n = 0, p = q = 1, l_1 = \iota_{\{0\}}, F_1 = \bar{\tau}I$ , with  $\bar{\tau} \in (0, +\infty)$ ,  $R_1 = \bar{\rho}I$  with  $\bar{\rho} \in (0, +\infty)$ ,  $E_1$  and  $G_1$  are finite dimensional spaces and  $\lambda_n \equiv 1, n \in \mathbb{N}$ , Algorithm (5.19) extends the algorithms in [16-17] which were developed in a deterministic setting.

- (iii) If  $\forall k \in \{1, \dots, q\}, B_k = \tilde{D}_k$ ,  $p = 1$  and  $\lambda_n \equiv 1, n \in \mathbb{N}$ , Algorithm (5.7) extends the deterministic approaches in [18] by introducing some random sweeping of the coordinates in the absence of errors.
- (iv) Theorem 3.2 extends the corresponding results Theorem 5 of Radu Ioan, Ernő Robert and Christopher [1] from a nonexpansive mapping to a quasinonexpansive mapping.
- (v) In Theorem 2.1, Corollary 2.3, Theorem 3.2 and Theorem 3.5, if  $\alpha_n = 0, \forall n \in \mathbb{N}$ , we can obtain Theorem 2.5, Corollary 2.7, Theorem 3.2 and Corollary 3.8 of Patrick, Combettes and Jean-Christophe [2] in the absence of errors.

## Acknowledgements

This work was supported by the National Natural Science Foundation of China (11131006, 41390450, 91330204, 11401293), the National Basic Research Program of China (2013CB 329404), the Natural Science Foundations of Jiangxi Province (CA201107114, 20114BAB 201004).

## References

- [1] R. I. Boţ, E. R. Csetnek and C. Hendrich, Inertial Douglas-Rachford splitting for monotone inclusion problems, *Appl. Math. Comput.*, vol. 256, pp. 472-487, 2015.
- [2] P. L. Combettes and J. C. Pesquet, Stochastic Quasi-Fejér Block-Coordinate Fixed Point Iterations with Random Sweeping, 2014, [http://www.optimization-online.org/DB\\_HTML/2014/04/4333.html](http://www.optimization-online.org/DB_HTML/2014/04/4333.html).
- [3] J. C. Pesquet and A. Repetti, A Class of Randomized Primal-Dual Algorithms for Distributed Optimization, 2014, <http://arxiv.org/abs/1406.6404v3>.
- [4] R. I. Boţ, and C. Hendrich, A Douglas-Rachford type primal-dual method for solving inclusions with mixtures of composite and parallel-sum type monotone operators. *SIAM J. Optim.*, 23(4):2541-2565, Dec. 2013.
- [5] L. M. Briceño-Arias, A Douglas-Rachford splitting method for solving equilibrium problems. *Nonlinear Anal.*, 75(16):6053-6059, Nov. 2012.

- [6] P. L. Combettes, Systems of structured monotone inclusions: duality, algorithms, and applications. *SIAM J. Optim.*, 23(4):2420-2447, Dec. 2013.
- [7] P. L. Combettes and J. C. Pesquet, Primal-dual splitting algorithm for solving inclusions with mixtures of composite, Lipschitzian, and parallel-sum type monotone operators. *Set-Valued Var. Anal.*, 20(2):307-330, June 2012.
- [8] J.C. Pesquet and N. Pustelnik, A parallel inertial proximal optimization method. *Pac. J. Optim.*, 8(2):273-305, Apr. 2012.
- [9] H. Raguet, J. Fadili, , and G. Peyré. A generalized forward-backward splitting. *SIAM J. Imaging Sci.*, 6(3):1199-1226, 2013.
- [10] P. L. Combettes. Systems of structured monotone inclusions: duality, algorithms, and applications. *SIAM J. Optim.*, 23(4):2420-2447, Dec. 2013.
- [11] A. Chambolle and T. Pock. A first-order primal-dual algorithm for convex problems with applications to imaging. *J. Math. Imaging Vision*, 40(1):120-145, 2011.
- [12] L. Condat. A primal-dual splitting method for convex optimization involving Lipschitzian, proximable and linear composite terms. *J. Optim. Theory Appl.*, 158(2):460-479, Aug. 2013.
- [13] E. Esser, X. Zhang, and T. Chan. A general framework for a class of first order primal-dual algorithms for convex optimization in imaging science. *SIAM J. Imaging Sci.*, 3(4):1015-1046, 2010.
- [14] B. He and X. Yuan. Convergence analysis of primal-dual algorithms for a saddle-point problem: from contraction perspective. *SIAM J. Imaging Sci.*, 5(1):119-149, 2012.
- [15] B. C. Vũ. A splitting algorithm for dual monotone inclusions involving cocoercive operators. *Adv. Comput. Math.*, 38(3):667-681, Apr. 2013.
- [16] P. Chen, J. Huang, and X. Zhang. A primal-dual fixed point algorithm for convex separable minimization with applications to image restoration. *Inverse Problems*, 29(2):025011, 2013.

- [17] I. Loris and C. Verhoeven. On a generalization of the iterative soft-thresholding algorithm for the case of non-separable penalty. *Inverse Problems*, 27(12):125007, 2011.
- [18] D. A. Lorenz and T. Pock. An Inertial Forward-Backward Algorithm for Monotone Inclusions. *J Math Imaging Vis* DOI 10.1007/s10851-014-0523-2, 2014.
- [19] N.Ogura, and I. Yamada, Non-strictly convex minimization over the fixed point set of an asymptotically shrinking nonexpansive mapping. *Numer. Funct. Anal. Optim.*, vol. 23 (1-2), pp. 113-137, 2002.
- [20] H. H. Bauschke and P. L. Combettes, A weak-to-strong convergence principle for Fejér-monotone methods in Hilbert spaces, *Math. Oper. Res.*, vol. 26, pp. 248-264, 2001.
- [21] H. H. Bauschke and P. L. Combettes, *Convex Analysis and Monotone Operator Theory in Hilbert Spaces*. Springer, New York, 2011.
- [22] P. L. Combettes, Fejér monotonicity in convex optimization, in: *Encyclopedia of Optimization*, 2nd ed. (C. A. Floudas and P. M. Pardalos, eds.), pp. 1016-1024. Springer, New York, 2009.
- [23] I. I. Eremin and L. D. Popov, Fejér processes in theory and practice: Recent results, *Russian Math. (Iz. VUZ)*, vol. 53, pp. 36-55, 2009.
- [24] E. Raik, Fejér type methods in Hilbert space, *Eesti NSV Tead. Akad. Toimetised Füü.-Mat.s*, vol. 16, pp. 286-293, 1967.
- [25] Yu. M. Ermole'v, On the method of generalized stochastic gradients and quasi-Fejer sequences, *Cybernetics*, vol. 5, pp. 208-220, 1969.
- [26] Yu. M. Ermol'ev, On convergence of random Fejér sequences, *Cybernetics*, vol. 7, pp. 655-656, 1971.
- [27] Yu. M. Ermol'ev and A. D. Tuniev, Random Fejér and quasi-Fejér sequences, *Theory of Optimal SolutionsC Akademiya Nauk Ukrainskoi SSR Kiev*, vol. 2, pp.



- 76-83, 1968; translated in: American Mathematical Society Selected Translations in Mathematical Statistics and Probability, vol. 13, pp. 143-148, 1973.
- [28] S. R. Becker and P. L. Combettes. An algorithm for splitting parallel sums of linearly composed monotone operators, with applications to signal recovery. *J. Nonlinear Convex Anal.*, 15(1):137-159, Jan. 2014.
  - [29] C. Couprie, L. Grady, L. Najman, J.-C. Pesquet, and H. Talbot. Dual constrained TV-based regularization on graphs. *SIAM J. Imaging Sci.*, 6:1246-1273, 2013.
  - [30] S. Harizanov, J.-C. Pesquet, and G. Steidl. Epigraphical projection for solving least squares Anscombe transformed constrained optimization problems. In A. Kuijper et al., editor, 4th International Conference on Scale-Space and Variational Methods in Computer Vision, volume 7893 of Lecture Notes in Computer Science, pages 125-136, Schloss Seggau, Leibnitz, Austria, 2-6 June 2013. Springer-Verlag, Berlin.
  - [31] A. Jezierska, E. Chouzenoux, J.-C. Pesquet, and H. Talbot. A primal-dual proximal splitting approach for restoring data corrupted with Poisson-Gaussian noise. In *Proc. Int. Conf. Acoust., Speech Signal Process.*, pages 1085-1088, Kyoto, Japan, 25-30 Mar. 2012.
  - [32] N. Pustelnik, P. Borgnat, and P. Flandrin. Empirical Mode Decomposition revisited by multicomponent nonsmooth convex optimization. *Signal Process.*, 102:313-331, Sept. 2014.
  - [33] A. Repetti, E. Chouzenoux, and J.-C. Pesquet. A penalized weighted least squares approach for restoring data corrupted with signal-dependent noise. In *Proc. Eur. Sig. and Image Proc. Conference*, pages 1553-1557, Bucharest, Romania, 27-31 Aug. 2012.
  - [34] T. Teuber, G. Steidl, and R.-H. Chan. Minimization and parameter estimation for seminorm regularization models with  $I$ -divergence constraints. *Inverse Problems*, 29:035007, Mar. 2013.
  - [35] Polyak, B.T.: Some methods of speeding up the convergence of iteration methods. *U.S.S.R. Comput. Math. Math. Phys.* 4(5), 1-17 (1964)

- [36] Eckstein, J., Bertsekas, D.P.: On the Douglas-CRachford splitting method and the proximal point algorithm for maximal monotone operators. *Math. Program.* 55, 293-318 (1992)
- [37] Nesterov, Y.: Introductory lectures on convex optimization: a basic course. In: *Applied Optimization*, vol. 87. Kluwer Academic Publishers, Boston, MA (2004)
- [38] A. Auslender, Méthodes numériques pour la décomposition et la minimisation de fonctions non différentiables, *Numer. Math.*, vol. 18, pp. 213-223, 1971/72.
- [39] J. C  a, *Optimisation: Th  orie et Algorithmes*, Dunod, Paris, 1971.
- [40] J. M. Ortega and W. C. Rheinboldt, *Iterative Solution of Nonlinear Equations in Several Variables*, Academic Press, New York, 1970.
- [41] Y. Nesterov, A method for solving the convex programming problem with convergence rate  $O(\frac{1}{k^2})$ , *Dokl. Akad. Nauk SSSR*, 269(3), 543-547, 1983.